

MÁXIMOS Y MINIMOS

Marco Antonio Cruz Chávez

UAEM Av. Universidad 1001 Col. Chamilpa C.P. 62210 Cuernavaca Morelos, México
Agosto 18 del 2000

00334858@academ01.mor.itesm.mx

Abstract. En este trabajo se presentan algunos métodos de optimización clásica relacionados con los máximos y mínimos de funciones no lineales. También se muestra una explicación del método de optimización simplex para funciones lineales por el cual se pueden obtener los máximos y mínimos de una función restringida.

1 INTRODUCCIÓN.

Desde la década de los 60 la programación lineal (PL) ha sido aplicada en diversas áreas de la vida como por ejemplo: sistemas militares, agrícolas, económicos, de transporte y de salud. La PL ofrece bases importantes en el desarrollo de métodos de solución de otras técnicas de la Investigación de operaciones, como lo son la programación entera, la estocástica y la no lineal [Taha 1991]. La PL juega un papel muy importante en el estudio de los problemas continuos de optimización considerados como la frontera de los problemas de optimización combinatoria, ya que en los continuos se tienen las características necesarias para que sean considerados dentro del tipo combinatorio [Papadimitriou and Steiglitz, 1982]: Un problema de optimización combinatoria siempre se le involucra un conjunto de instancias, donde cada una de ellas cuenta con un conjunto finito de posibles soluciones (característica imprescindible de los problemas continuos).

Por otra parte la teoría de optimización clásica se usa para la obtención de los máximos y mínimos de funciones no lineales restringidas y no restringidas, en los que se hace uso del calculo diferencial.

2 MAXIMOS Y MINIMOS

Mínimo (fuerte): Un punto extremo X_0 de una función $f(X_0)$ define un mínimo de la función si $f(X_0+h) > f(X_0)$, donde X_0 es cualquier punto de la función y h en valor absoluto es suficientemente pequeña.

Máximo (fuerte): Un punto extremo X_0 de una función $f(X_0)$ define un máximo de la función si $f(X_0+h) < f(X_0)$, donde X_0 es cualquier punto de la función y h en valor absoluto es suficientemente pequeña.

Una función puede contener varios máximos y mínimos, identificados por los puntos extremos de la función. En la figura 1 se puede observar esto, los puntos x_1 , x_3 y x_6 son máximos, de la figura notamos que $f(x_6)$ es el mayor que $f(x_1)$ y $f(x_3)$, a este punto se le conoce como máximo global de la función y a los restantes como máximos locales. Lo mismo se puede ver para los mínimos, en los que también existe un mínimo global $f(x_2)$ y un mínimo local $f(x_4)$. Como es de lógico, solo puede existir un solo global y posiblemente varios locales.

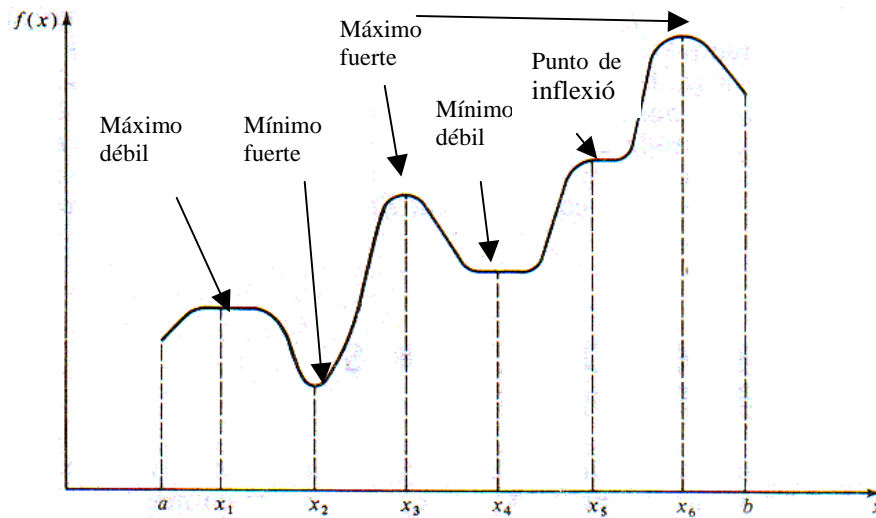


Fig. 1. Representación de máximos y mínimos en una función con una sola variable [Taha 1991].

Una condición necesaria pero no suficiente para que X_0 sea un punto extremo, es que para una función con mas de una variable, el gradiente $\nabla f(X_0) = 0$. Si es cierto esto entonces X_0 será conocido como punto estacionario.

Una condición suficiente para que un punto estacionario sea extremo es que la matriz Hessiana \mathbf{H} obtenida en X_0 del sistema de ecuaciones sea positiva cuando X_0 es un punto extremo de mínimo. Y negativa cuando X_0 es un punto extremo de máximo.

Un máximo débil implica un numero finito de máximos alternativos (ver figura 1) y se define como X_0 es un máximo débil, si $f(X_0 + h) \leq f(X_0)$. Un análisis similar es para los mínimos débiles.

Un punto de inflexión se encuentra cuando la evaluación del gradiente da cero y no es un extremo, esto es, se debe de cumplir la condición de la matriz Hessiana.

3 TÉCNICAS DE OPTIMIZACION CLÁSICA.

3.1 Método de Newton

El encontrar la solución a un sistema de ecuaciones no lineal es mucho más difícil que el de un sistema lineal. El método de Newton es un procedimiento iterativo y permite la linealización de un sistema de ecuaciones no lineal, para posteriormente darle solución por cualquier método numérico de ecuaciones lineales simultáneas, este forma parte de los métodos conocidos como métodos de gradiente.

Un sistema de n ecuaciones con n incógnitas (x_1, x_2, \dots, x_n) , se conoce como no lineal, si una o más de estas no es lineal.

De manera general, la solución de un sistema de n ecuaciones no lineal aplicando el método de Newton [Conte y Boor, 1987] se plantea como sigue:

$$\begin{bmatrix} \frac{\partial f_1(x, y, z, \dots)}{\partial x} & \frac{\partial f_1(x, y, z, \dots)}{\partial y} & \frac{\partial f_1(x, y, z, \dots)}{\partial z} & \dots \\ \frac{\partial f_2(x, y, z, \dots)}{\partial x} & \frac{\partial f_2(x, y, z, \dots)}{\partial y} & \frac{\partial f_2(x, y, z, \dots)}{\partial z} & \dots \\ \frac{\partial f_3(x, y, z, \dots)}{\partial x} & \frac{\partial f_3(x, y, z, \dots)}{\partial y} & \frac{\partial f_3(x, y, z, \dots)}{\partial z} & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n(x, y, z, \dots)}{\partial x} & \frac{\partial f_n(x, y, z, \dots)}{\partial y} & \frac{\partial f_n(x, y, z, \dots)}{\partial z} & \dots \end{bmatrix} \begin{bmatrix} \Delta x_i \\ \Delta y_i \\ \Delta z_i \\ \vdots \end{bmatrix} \quad (1)$$

evaluado en (x_i, y_i, z_i, \dots) y resolviendo se tiene:

$$x_{i+1} = x_i + \Delta x_i, \quad y_{i+1} = y_i + \Delta y_i, \quad z_{i+1} = z_i + \Delta z_i, \dots$$

Sea la expresión (2) la suma de todas las ecuaciones no lineales que representan el sistema en estudio

$$f_{2n}(\underline{X}^{k+1}) = D_n(\underline{X}^{k+1}) + E_n(\underline{X}^{k+1}) \quad (2)$$

La iteración de Newton se podrá obtener para el conjunto de ecuaciones a partir de (2), usando la serie de Taylor en una aproximación de primer orden

$$f_{2n}^{k+1}(\underline{X}_{i+1}) \approx f_{2n}^{k+1}(\underline{X}_i) + f'_{2n}{}^{k+1}(\underline{X}_i)(\underline{X}_{i+1} - \underline{X}_i) = 0 \quad (3)$$

Las ecuaciones (3) representan un conjunto de $2n$ ecuaciones lineales acopladas para el vector de incógnitas \underline{X} en el paso de tiempo $k+1$, en la iteración $i+1$. Los términos en derivadas darán lugar a la formación de la matriz Jacobiana:

$$f'_{2n^{k+1}}(\underline{X}_i) = \sum \frac{\nabla f_{2n^{k+1}}(\underline{X}_i)}{\nabla \underline{X}_i} \quad (4)$$

La solución del conjunto de ecuaciones lineales encontradas en cada iteración (matriz Jacobiana) será resuelta por cualquier método lineal de matrices hasta que los residuales sean menores a una tolerancia designada, muy próxima a cero, ecuación (3).

3.2 Método de derivadas restringidas (Jacobiano).

Se puede considerar como una generalización del método simplex para programación lineal. Considere el problema

$$\begin{aligned} \text{Minimizar } z &= f(\underline{X}) \\ \text{Sujeto a } g(\underline{X}) &= 0 \end{aligned}$$

$$\begin{aligned} \text{Donde } \underline{X} &= (x_1, x_2, \dots, x_n) \\ g &= (g_1, g_2, \dots, g_m)^T \end{aligned}$$

Las funciones $f(\underline{X})$ y $g_i(\underline{X})$, donde $i = 1, 2, \dots, m$, se suponen diferenciables y doblemente continuas. El utilizar derivadas restringidas es encontrar una expresión de forma cerrada para las primeras derivadas parciales de $f(\underline{X})$ en todos los puntos que satisfacen las restricciones $g(\underline{X}) = 0$. Estos puntos estacionarios entonces se sabe que son donde dichas derivadas parciales se anulan.

Por el teorema de Taylor, para los puntos $\underline{X} + \Delta \underline{X}$ en el entorno factible de \underline{X} , se deduce que

$$\begin{aligned} f(\underline{X} + \Delta \underline{X}) - f(\underline{X}) &= \nabla f(\underline{X})\Delta \underline{X} + 0(\Delta \underline{X}^2) \\ &\text{y} \\ g(\underline{X} + \Delta \underline{X}) - g(\underline{X}) &= \nabla g(\underline{X})\Delta \underline{X} + 0(\Delta \underline{X}^2) \end{aligned}$$

Cuando $\Delta x_j \rightarrow 0$ se tiene que

$$\frac{\partial f(\underline{X})}{\partial \underline{X}} = \nabla f(\underline{X})\Delta \underline{X} \quad \text{y} \quad \frac{\partial g(\underline{X})}{\partial \underline{X}} = \nabla g(\underline{X})\Delta \underline{X} \quad (5)$$

ya que $g(\underline{X}) = 0$, entonces $\frac{\partial g(\underline{X})}{\partial \underline{X}} = 0$, se deduce que,

$$\begin{aligned} \frac{\partial f(\underline{X})}{\partial \underline{X}} - \nabla f(\underline{X}) \frac{\partial \underline{X}}{\partial \underline{X}} &= 0 \\ \nabla g(\underline{X}) \frac{\partial \underline{X}}{\partial \underline{X}} &= 0 \end{aligned}$$

con esto se tienen m ecuaciones con n incógnitas, cuando $m < n$ tenemos, si definimos a \underline{X} ahora como:

$$\underline{X} = (\underline{Y}, \underline{Z})$$

Donde

$$Y = (y_1, y_2, \dots, y_m) \text{ y } Z = (z_1, z_2, \dots, z_{n-m})$$

Indican a las variables dependientes e independientes, respectivamente y que corresponden al vector X. Volviendo a escribir los vectores gradiente de f y g en términos de Y y Z se encuentra que

$$\nabla f(Y, Z) = (\nabla_y f, \nabla_z f)$$

$$\nabla g(Y, Z) = (\nabla_y g, \nabla_z g)$$

Donde

$$J = \nabla_y g = \begin{pmatrix} \nabla_y g_1 \\ \vdots \\ \nabla_y g_m \end{pmatrix}$$

$$C = \nabla_z g = \begin{pmatrix} \nabla_z g_1 \\ \vdots \\ \nabla_z g_m \end{pmatrix}$$

$J_{m \times m}$ se conoce como la matriz jacobiana y $C_{m \times n-m}$ es la matriz de control. Utilizando las definiciones anteriores en (4), se puede escribir el conjunto original de ecuaciones como

$$\partial f(Y, Z) = \nabla_y f \partial Y + \nabla_z f \partial Z \quad (6)$$

y

$$J \partial Y = - C \partial Z$$

Como J se supone que no es singular, existirá su inversa J^{-1} . Por lo tanto,

$$\partial Y = - J^{-1} C \partial Z$$

sustituyendo ∂Y en la ecuación (6) se obtiene a ∂f como una función de ∂Z

$$\partial f(Y, Z) = (\nabla_z f - \nabla_y f J^{-1} C) \partial Z$$

aplicando a esta ecuación la derivada restringida con respecto a al vector independiente Z, se tiene

$$\nabla_c f = \partial_c f(Y,Z) / \partial_c Z = \nabla_z f - \nabla_y f J^{-1} C$$

donde $\nabla_c f$ representa el vector gradiente restringido de f con respecto a Z. Por consiguiente f debe de ser nulo en los puntos estacionarios.

3.3 Método Lagrangiano.

Esta muy relacionado con el método jacobiano y se puede desarrollar a partir de aquel. Anteriormente se demostró que:

$$\partial f(Y,Z) = \nabla_y f \partial Y + \nabla_z f \partial Z \quad (7)$$

$$\partial g = J \partial Y + C \partial Z$$

Supongamos que ∂g no es igual a 0 entonces despejamos ∂Y

$$\partial Y = J^{-1} \partial g - J^{-1} C \partial Z$$

sustituyendo en la ecuación (6) tenemos

$$\partial f(Y,Z) = \nabla_y f J^{-1} \partial g + \nabla_c f \partial Z \quad (8)$$

donde

$$\nabla_c f = \nabla_z f - \nabla_y f J^{-1} C$$

Considerando que en el punto extremo (y para cualquier punto estacionario) $X_0 = (Y_0, Z_0)$, el gradiente restringido $\nabla_c f$ debe anularse, entonces de (8)

$$\partial f(Y_0, Z_0) = \nabla_{y_0} f J^{-1} \partial g(Y_0, Z_0)$$

o bien,

$$\partial f / \partial g = \nabla_{y_0} f J^{-1} \quad (9)$$

Esta relación (9) permite estudiar el efecto de variaciones pequeñas de g sobre el valor optimo de f evaluando la tasa de cambio de f con respecto a g. Estas tasas usualmente se conocen como coeficientes de sensibilidad.

De (9) si tomamos

$$\lambda = \nabla_{y_0} f J^{-1} = \partial f / \partial g$$

tendremos

$$\partial f - \lambda \partial g = 0 \quad (10)$$

Esta ecuación satisface las condiciones necesarias para los puntos estacionarios ya que la expresión para $\partial f / \partial g$ se calcula con $\nabla_c f = 0$. Otra representación mas conveniente de la ecuación (10) es tomando sus derivadas parciales con respecto a todas las x_j . Lo anterior da

$$\frac{\partial}{\partial x_j} (f - \lambda g) = 0 \quad j = 1, 2, \dots, n$$

Las ecuaciones resultantes junto con las de restricciones $g = 0$ proporcionan los valores factibles de X y λ que satisfacen las condiciones necesarias para puntos estacionarios. Este procedimiento define al método de Lagrange para identificar los puntos estacionarios de problemas de optimización con restricciones de igualdad.

4 EL METODO SIMPLEX

Para resolver un modelo de PL por medio del método simplex, este se deberá de poner en forma estándar aplicando los siguientes puntos:

1. Todas las restricciones son ecuaciones con un segundo miembro positivo.
2. Una restricción se convierte en ecuación adicionando una variable de holgura (o restando una variable de exceso) al primer miembro de la ecuación. El segundo miembro de la ecuación se puede hacer positivo, multiplicando ambos lados por -1 . Y la dirección de una desigualdad se invierte multiplicando ambos miembros de la ecuación por -1 .
3. Todas las variables son positivas.
4. En el caso de que una variable y_i cualquiera, sea irrestricta (no restringida), esta pudiera expresarse en términos de dos variables positivas mediante el uso de la sustitución. $y_i = y'_i - y''_i$ donde $y'_i, y''_i \geq 0$.
5. La función objetivo puede ser de maximización o de minimización.

Un ejemplo sencillo para poner un modelo de PL en forma estándar es el siguiente:

Modelo de PL.

$$\text{Minimizar } z = 2x_1 + 3x_2$$

sujeto a

$$x_1 + x_2 = 10$$

$$\begin{aligned}
-2x_1 + 3x_2 &\leq -5 \\
7x_1 - 4x_2 &\leq 6 \\
x_1 &\text{ irrestricta} \\
x_2 &\geq 0
\end{aligned}$$

Haciendo uso de los puntos anteriores obtenemos el siguiente modelo PL en forma estándar:

$$\text{Minimizar } z = 2x'_1 - 2x''_1 + 3x_2$$

sujeto a

$$\begin{aligned}
x'_1 - x''_1 + x_2 &= 10 \\
2x'_1 - 2x''_1 + 3x_2 - s_2 &= -5 \\
7x'_1 - 7x''_1 - 4x_2 + s_3 &= 6 \\
x_1, x''_1, x_2, x_3, s_2, s_3 &\geq 0
\end{aligned}$$

La solución óptima a un modelo PL se encontrará siempre en algún punto extremo del espacio de soluciones formado por el conjunto de restricciones, a cada punto extremo se le conoce como (solución básica factible). Simplex emplea un proceso iterativo que inicia en un punto extremo factible (por lo general el origen) considerada como la solución inicial y va recorriendo de un punto factible a otro punto factible, hasta que encuentra el punto óptimo, punto en el cual (en el caso de maximización), el valor de la función objetivo sea el máximo que se pueda obtener. Para desplazarse a un punto factible se emplea una iteración en el método.

Los pasos del algoritmo simple son los siguientes:

Paso 0: después de poner el modelo PL en forma estándar, se determina una solución básica factible inicial, inicializando un número W de Variables no Básicas ($W = \text{incógnitas} - \text{ecuaciones}$) a un valor igual a cero.

Paso 1: seleccionar una variable de entrada de las variables no básicas actuales (cero), que al incrementarse esta arriba de cero, contribuye al incremento de la función objetivo (maximización), si no existe una menor o igual a cero, la solución actual es óptima. De lo contrario hacer el paso dos.

Paso 2: seleccionar una variable de salida del conjunto de variables básicas actuales y hágala cero (vuélvala no básica) para que la nueva básica (variable de entrada) seleccionada en el paso uno tome su lugar.

Paso 3: determinar la nueva solución básica haciendo que la variable de entrada sea básica y la variable de salida sea no básica. Continuar en el paso uno para la siguiente iteración.

Variable no básica: son las variables del modelo PL que se hacen igual a cero al inicio de la primera iteración.

Variable básica: son las variables del modelo PL que son diferente de cero al inicio de la primera iteración.

Para entender mas claro este algoritmo, consideremos un ejemplo muy sencillo: Modelo PL de Reddy Mikks en su forma estándar, para mas detalles ver referencia [Taha 1991]:

$$\begin{array}{rcl}
 z - 3x_1 - 2x_2 & = & 0 \quad \text{Función objetivo} \\
 x_1 + 2x_2 + s_1 & = & 6 \\
 2x_2 + x_1 + s_2 & = & 8 \\
 -x_1 + x_1 + s_3 & = & 1 \\
 x_1 + s_4 & = & 2
 \end{array} \left. \vphantom{\begin{array}{rcl}} \right\} \text{Restricciones}$$

El numero de variables no básicas (tomaran el valor inicial de cero) será $W = 6 - 4 = 2$.

La solución inicial básica factible, si $x_1 = x_2 = 0$, es $s_1 = 6$, $s_2 = 8$, $s_3 = 1$, $s_4 = 2$ y el valor de $z = 0$. Presentando esto en forma tabular se tiene.

básica	z	x_1	x_2	s_1	s_2	s_3	s_4	solución
z	1	-3	-2	0	0	0	0	0
s_1	0	1	2	1	0	0	0	6
s_2	0	2	1	0	1	0	0	8
s_3	0	-1	1	0	0	1	0	1
s_4	0	0	1	0	0	0	1	2

La columna llamada básica, indica a las variables básicas actuales s_1 , s_2 , s_3 y s_4 . De forma implícita las variables no básicas son x_1 y x_2 pues no se encuentran en la columna básica.

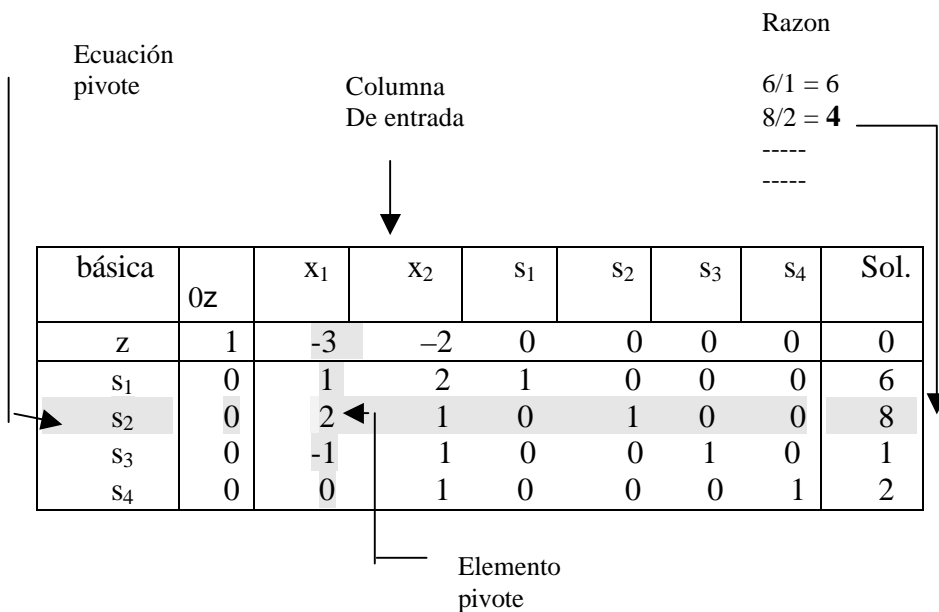
Para detectar cuando se encuentra la solución optima (la mejor) hacemos uso de:

Condición de optimidad: en el caso de la maximización, si todas las variables no básicas tienen coeficientes positivos en la función objetivo, se ha encontrado la solución optima. En caso contrario se selecciona a la variable con coeficiente más negativo como la variable de entrada.

Para detectar a la variable de salida en caso de que no se haya encontrado la solución optima hacemos uso de:

Condición de factibilidad: se obtiene la mínima razón = solución/(coeficiente positivo), donde el coeficiente positivo es asociado a la variable de entrada, y no se toman en cuenta valores menores o iguales a cero.

De esta manera con estas dos condiciones podemos obtener a la columna de entrada y a la ecuación pivote que señalara a la variable básica de salida. Esto se ve mas claro en la siguiente tabla.



Una vez que se identifican a las variables de entrada y salida la nueva solución básica (siguiente iteración) se encuentra al aplicar el método de Gauss-Jordan que involucra los siguientes pasos:

1. Nueva ecuación pivote = ecuación pivote anterior/elemento pivote.
2. Todas las otras ecuaciones incluyendo z.
 Nueva ecuación = ecuación anterior - (su coeficiente de la columna de entrada) x (nueva ecuación pivote).

Al aplicar esto a la tabla anterior se tiene lo siguiente:

Razon

$2/(3/2) = 4/3$
 $4/(1/2) = 8$
 $5/(3/2) = 10/3$
 $2/1 = 2$

básica	1z	x ₁	x ₂	s ₁	s ₂	s ₃	s ₄	solución
z	1	0	-1/2	0	3/2	0	0	12
s ₁	0	0	3/2	1	-1/2	0	0	2
x ₁	0	1	1/2	0	1/2	0	0	4
s ₃	0	0	3/2	0	1/2	1	0	5
s ₄	0	0	1	0	0	0	1	2

La nueva solución será $x_1 = 4$, $x_2 = 0$ y $z = 12$. Para encontrar la columna de entrada y el renglón pivote, se repite a partir de las condiciones de optimalidad y factibilidad hasta que se encuentre la solución óptima (cuando se cumple la condición de optimalidad).

5 LA PROGRAMACIÓN NO LINEAL: METODOS DE GRADIENTE.

El método del gradiente se utiliza para la optimización de funciones que son dos veces diferenciables. La idea general para generar puntos sucesivos, iniciando de un punto inicial dado, en la dirección del incremento más rápido (maximización) de la función. La técnica es conocida como método de gradiente debido a que el gradiente de la función en un punto es indicativo de la rapidez del incremento.

5.1 Método Newton-Raphson

Uno de los métodos de gradiente es el Newton-Raphson, este método está basado en la resolución de ecuaciones simultáneas que representan la condición necesaria para optimalidad, llamada $\nabla f(\mathbf{X}) = 0$.

La desventaja de utilizar la condición necesaria $\nabla f(\mathbf{X}) = 0$ para determinar los puntos estacionarios es la dificultad para resolver las ecuaciones simultáneas resultantes numéricamente. El método Newton-Raphson es un procedimiento iterativo para resolver ecuaciones simultáneas no lineales.

Considere las ecuaciones simultáneas:

$$f_i(\mathbf{X}) = 0, \quad i = 1, 2, \dots, m$$

donde X^k es un punto dado. Por expansión de Taylor

$$f_i(X) \cong f_i(X^k) + \nabla f_i(X^k)(X - X^k), \quad i = 1, 2, \dots, m$$

Así, la ecuación original puede aproximarse a

$$f_i(X^k) + \nabla f_i(X^k)(X - X^k) = 0, \quad i = 1, 2, \dots, m$$

Estas ecuaciones pueden escribirse en notación matricial como

$$\mathbf{A}_k + \mathbf{B}_k(\mathbf{X} - \mathbf{X}^k) = 0$$

Tomando en cuenta la condición de que toda $f_i(\mathbf{X})$ es independiente, \mathbf{B}_k es necesariamente no singular. Así, la ecuación anterior resulta

$$\mathbf{X} = \mathbf{X}^k - \mathbf{B}_k^{-1}\mathbf{A}_k$$

La idea del método es iniciar de un punto inicial \mathbf{X}^0 . Utilizando la ecuación anterior, un nuevo punto \mathbf{X}^{k+1} puede determinarse de \mathbf{X}^k . El procedimiento termina con \mathbf{X}^m como solución cuando $\mathbf{X}^m \cong \mathbf{X}^{m-1}$.

Una interpretación geométrica de método se muestra con una función de una sola variable en la figura 2. La relación entre x^k y x^{k+1} para una función de una sola variable $f(x)$ se reduce a

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}$$

o

$$f'(x^k) = \frac{f(x^k)}{x^k - x^{k+1}}$$

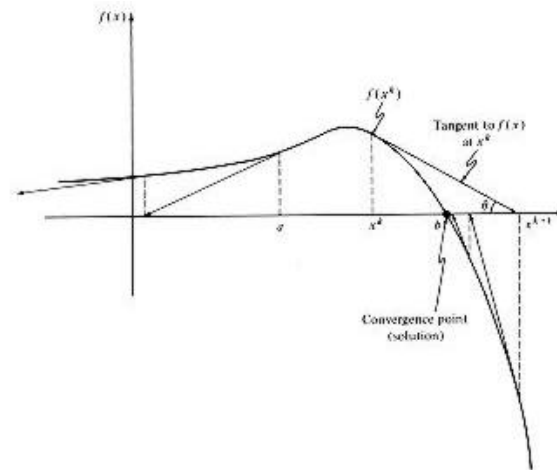


Fig. 2. Representación grafica de Newton Raphson.

Analizando la figura 2 se puede ver que x^{k+1} se determinan de la pendiente de $f(x)$ en x^k , donde $\tan \theta = f'(x^k)$.

Un problema con este método es que la convergencia no esta garantizada siempre a menos que la función se comporte bien. En la figura, si el punto inicial x_0 es a el método divergirá. No existe una manera fácil para localizar un "buen" punto inicial x_0 . Tal vez una solución para este problema es usar prueba y error.

5.2 Método de ascenso pronunciado (steepest ascent)

La terminación del método del gradiente se realiza en el punto donde el gradiente de $f(\mathbf{X})$ vuelve nulo. Esta es solamente una condición necesaria para la optimalidad. Así se enfatiza que la optimalidad no puede verificarse al menos que se conozca a priori que $f(\mathbf{X})$ es cóncava o convexa.

Suponga que $f(\mathbf{X})$ es maximizada. Tomemos a \mathbf{X}^0 como el punto inicial en el cual el procedimiento inicia y define a $\nabla f(\mathbf{X}^k)$ como el gradiente de f en el k -ésimo punto de \mathbf{X}^k . La idea del método es determinar el PATH p a través del cual df/dp es maximizada en un punto dado. Este resultado es alcanzado si los puntos sucesivos \mathbf{X}^k y \mathbf{X}^{k+1} se seleccionan tales que

$$\mathbf{X}^{k+1} = \mathbf{X}^k + r^k \nabla f(\mathbf{X}^k)$$

donde r^k es un parámetro llamado **tamaño de paso óptimo**.

El parámetro r^k se determina tal que \mathbf{X}^{k+1} de cómo resultado una gran mejora de f . En otras palabras, si la función $h(r)$ se define como

$$h(r) = f(\mathbf{X}^k + r \nabla f(\mathbf{X}^k))$$

r^k es el valor de r maximizando

El procedimiento propuesto termina cuando dos puntos de prueba sucesivos \mathbf{X}^k y \mathbf{X}^{k+1} son aproximadamente iguales. Esto es equivalente a tener

$$r^k \nabla f(\mathbf{X}^k) \cong 0$$

Bajo el supuesto de que $r^k \neq 0$, el cual siempre será verdadero al menos que \mathbf{X}_0 se convierta en óptimo de $f(\mathbf{X})$, esto es equivalente a la condición necesaria $\nabla f(\mathbf{X}^k) = 0$

Ejemplo: Considere la función maximizada

$$f(x_1, x_2) = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2$$

$f(x_1, x_2)$ es una función cuadrática cuyo óptimo absoluto ocurre en $(x_1^*, x_2^*) = (1/3, 4/3)$. Esta muestra la manera en que el problema se resuelve con el método de ascenso pronunciado. La figura 3 muestra los puntos sucesivos. Los gradientes en cualquiera de los dos puntos sucesivos son necesariamente ortogonales (perpendiculares).

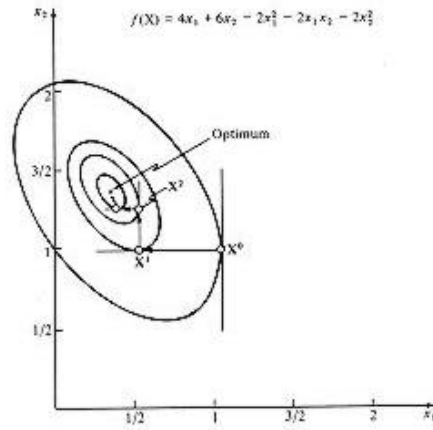


Fig. 3. Método de ascenso pronunciado.

Dado el punto inicial por $\mathbf{X}^0 = (1,1)$. Ahora

$$\nabla f(\mathbf{X}) = (4 - 4x_1 - 2x_2, 6 - 2x_1 - 4x_2)$$

Primera iteración

$$\nabla f(\mathbf{X}^0) = (-2,0)$$

El siguiente punto \mathbf{X}^1 se obtiene considerando

$$\mathbf{X} = (1,1) + r(-2,0) = (1 - 2r,1)$$

Así

$$h(r) = f(1 - 2r,1) = -2(1 - 2r)^2 + 2(1 - 2r) + 4$$

El tamaño del paso óptimo produce el valor máximo de $h(r)$ es $r^1 = 1/4$. Esto da como resultado $\mathbf{X}^1 = (1/2,1)$

Segunda iteración

$$\nabla f(\mathbf{X}^1) = (0,1)$$

Considerar

$$\mathbf{X} = (1/2,1) + r(0,1) = (1/2,1 + r)$$

Así

$$h(r) = -2(1 + r)^2 + 5(1 + r) + 3/2$$

Esto da como resultado

$$r^2 = 1/4 \text{ o } \mathbf{X}^2 = (1/2,5/4)$$

Tercera iteración

$$\nabla f(\mathbf{X}^2) = (-1/2, 0)$$

Considerar

$$\mathbf{X} = \left(\frac{1}{2}, \frac{5}{4}\right) + r \left(-\frac{1}{2}, 0\right) = \left(\frac{1-r}{2}, \frac{5}{4}\right)$$

Así

$$h(r) = -(1/2)(1-r)^2 + (3/4)(1-r) + 35/8$$

Esto da como resultado

$$r^3 = 1/4 \text{ o } \mathbf{X}^3 = (3/8, 5/4)$$

Cuarta iteración

$$\nabla f(\mathbf{X}^3) = (0, 1/4)$$

Considerar

$$\mathbf{X} = \left(\frac{3}{8}, \frac{5}{4}\right) + r \left(0, \frac{1}{4}\right) = \left(\frac{3}{8}, \frac{5+r}{4}\right)$$

Así

$$h(r) = -(1/8)(5+r)^2 + (21/16)(5+r) + 39/32$$

Da como resultado

$$r^4 = 1/4 \text{ o } \mathbf{X}^4 = (3/8, 21/16)$$

Quinta iteración

$$\nabla f(\mathbf{X}^4) = (-1/8, 0)$$

Considerar

$$\mathbf{X} = \left(\frac{3}{8}, \frac{21}{16}\right) + r \left(-\frac{1}{8}, 0\right) = \left(\frac{3-r}{8}, \frac{21}{16}\right)$$

sí

$$h(r) = -(1/32)(3-r)^2 + (11/64)(3-r) + 567/128$$

Da como resultado

$$r^5 = 1/4 \text{ o } \mathbf{X}^5 = (11/32, 21/16)$$

Sexta iteración

$$\nabla f(\mathbf{X}^5) = (0, 1/16)$$

Desde que $\nabla f(\mathbf{X}^5) \cong 0$, el proceso puede ser terminado en este punto. El punto máximo aproximado esta dado por $\nabla f(\mathbf{X}^5) = (0.3437, 1.3125)$. Es necesario observar que el óptimo exacto es $\mathbf{X}^* = (0.3333, 1.3333)$

6 CONCLUSIONES.

Se presento la teoría clásica de optimización para encontrar máximos y mínimos de los problemas no lineales restringidos. Y la conclusión es que no es adecuada para fines de calculo. En el caso del método simplex para sistemas lineales las condiciones de optimidad y factibilidad garantizan, partiendo de un punto extremo factible (solución básica), el poder mejorar el valor de la función objetivo hasta llegar al optimo en cada iteración.

Referencias.

1. Conte S.D. y Boor C; Análisis Numerico; McGraw-Hill, Mexico D.F., 1987.
2. C.H. Papadimitriou and K. Steiglitz, "Combinatorial optimization: algorithms and complexity", Prentice Hall Inc., USA. ISBN 0-13-152462-3, 496 pp., 1982.
3. H. A. Taha, Investigación de Operaciones, Alfaomega, Mexico D.F., ISBN 0-02-418940-5, 989 pp., 1991.